TISSUE CHARACTERIZATION AND DETECTION OF DYSPLASIA USING SCATTERED LIGHT

Fernand S. Cohen, Ezgi Taslidere, and Dilip S. Hari

Department of Electrical and Computer Engineering
Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104, USA

ABSTRACT

In this paper, the structural parameters of dysplasia formation in the epithelial tissue are estimated using a stochastic decomposition algorithm (SDM) by means of scattered light. We extract texture parameters obtained from the decomposition that capture the signature of dysplasia formation. These parameters include the number and mean energy of coherent scatterers; deviation from Rayleigh scattering; average energy of diffuse scatterers; and normalized correlation coefficient. The tests are performed on simulations, and tissue-mimicking phantom data. The simulations are based on the light scattered from the cells with varying parameters such as, index of refraction, number of cells, and size of cells. The obtained results demonstrate the proof-of-concept in being able to differentiate between tissue structures that give rise to changes in cell morphology as well as other physical properties such as change in index of refraction. Fusing all the estimated parameter set together results in the differentiation performance (Az value) up to 1(perfect detection) for simulated data, and Az>0.927 for the phantom data.

1. INTRODUCTION

About 90% of all cancers arise from epithelial cells that cover surface of the body and lining of the internal organs [1]. Dysplasia is a precursor tissue change prior to epithelial cancer, which can be reversible if it is detected at its very early stages. Dysplasia is invisible to eye, and can only be detected using biopsy [1]. Investigators are working on an alternative for biopsy to detect precancerous conditions precisely [2]. Since the dysplasia formation occurs in a size less than 1mm, optical techniques are a good choice for detection. Moreover, optical techniques do not necessitate tissue removal, and analysis can be made in real time.

The detailed spatial intensity distribution of light scattered by an individual particle is a complex function of the particle's size, shape and orientation with respect to the wavelength as well as the incident illumination [4]. Hence, the scattered light provides an objective measure of epithelial nuclear enlargement and crowding which are the most significant characteristics of dysplasia and early cancer [1]. Beyond the current state-of-the-art reflectance and fluorescence measurements, our method enables greater sensitivity by imaging the tissue area under investigation through the use of scanning with a detector array in order to construct a 2-D "image" of the cellular structure (A and B scans).

Our aim is to develop a reliable system for detection of epithelium dysplasia formation from scattered light by tracking down the characteristics of dysplasia as it develops in epithelium. We simulate the light scattered from the cells using a stochastic simulator based on Mie scattering. We extract tissue related parameters using a tissue characterization method, namely the Stochastic Decomposition Method [3], which separate the specular from the diffused scatterers. We then use these estimated parameters to differentiate between the diseased and normal cells. We show the results on simulations and tissue-mimicking phantom data. As indicated by the results, some of these parameters turn out to be very strong parameters to discriminate between the normal and diseased cells.

The paper is organized as follows. Section 2 presents dysplasia formation and details of scattered light structure. Section 3 presents the details of the simulation of the scattered light from the cells. The Stochastic Decomposition Method used for tissue related parameter extraction and the extracted features are described in Section 4. The results for the performance of these features for the discrimination of diseased cells and normal cells are given in Section 5. We conclude the paper in Section 6.

2. DYSPLASIA FORMATION AND SCATTERING EVENTS OF EPITHELIAL TISSUE

Dysplasia is a term that refers to a precancerous condition. Most of the time, malignant neoplastic changes follow pre-existing dysplastic changes. Removal of adverse environmental stimulus leads to restoration of normal cell growth pattern, hence dysplasia is reversible if it can be detected at its very early stages. Dysplasia is recognized by alterations in the appearance of cells (cytology). As tissue becomes dysplastic, the nuclei enlarge and become crowded. Healthy tissue epithelial nuclei have a characteristic diameter 4-7μm and are arranged in neat rows. In dysplastic epithelium, the cells proliferate and the nuclei enlarge and appear darker when stained. Nuclei can be as large as 20 μm in height. At the same magnification, normal intestinal cells viewed are characterized by uniform nuclear size distribution where malignant cells have larger nuclei and more variation in nuclear size [5]. In [5], for the normal tissue sample, the average diameter was found to be 4.8μm, standard deviation of the sizes was 0.4μm, and for the cancerous tissue sample the corresponding values were 9.75μm, and 1.5μm.

When particles are large compared to the wavelength of the incident light, Mie scattering theory describes the scattered light [6]. Scattering of light in tissue mainly consists of two components: singly scattered and multiply scattered. Singly

scattered light, as predicted by Mie Theory, is not randomized and contains information about individual scatterers. On the other hand, multiply scattered light is thoroughly randomized because of the multiple scattering that it undergoes and thus does not contain any specific information about the nature of scatterers. However, diffusely scattered light from tissue contains information about its basic structures. In the epithelial tissue, observation of single light scattering events is difficult, since the epithelial tissue is a turbid medium where the light transport is dominated by multiple scattering [7] [8].

3. SIMULATION OF SINGLE SPHERE AND MULTI SPHERE SCATTERING

3.1. Single Sphere Scattering

The relationship between incident and scattered electric field components perpendicular and parallel to the scattering plane as observed in the "far-field" is described by the amplitude scattering matrix:

$$\begin{bmatrix} E_{\parallel s} \\ E_{\perp s} \end{bmatrix} = \frac{e^{i\mathbf{k}(r-z)}}{-i\mathbf{k}r} \begin{bmatrix} S_2 & S_3 \\ S_4 & S_1 \end{bmatrix} \begin{bmatrix} E_{\parallel i} \\ E_{\perp i} \end{bmatrix} \quad (1)$$

where $k=2\pi n/\lambda$ is the wavenumber, λ is the wavelength of the incident light, and n is the scatterer refractive index. The amplitude scattering matrix assumes a particularly simple form when the scattering particles are spherical; S_3 and S_4 are equal to zero. The intensity and polarizing properties of the field are described via a Stokes vector [15] which specifies the set of parameters associated with the phase and polarization of radiation. The relation between incident and Stokes parameters is written in general form as [14]:

$$\begin{bmatrix} I_s \\ Q_s \\ U_s \\ V_s \end{bmatrix} = \frac{1}{k^2 r^2} \begin{bmatrix} S_{11} & S_{12} & S_{13} & S_{14} \\ S_{21} & S_{22} & S_{23} & S_{24} \\ S_{31} & S_{32} & S_{33} & S_{34} \\ S_{41} & S_{42} & S_{43} & S_{44} \end{bmatrix} \begin{bmatrix} I_i \\ Q_i \\ U_i \\ V_i \end{bmatrix} \quad (2)$$

where I is the total intensity, Q is the degree of linear polarization, U and V characterize the phase. The elements of the Mueller scattering matrix that we need in our study are described in terms of the amplitude scattering matrix elements as [14]:

$$S_{11} = \frac{1}{2}\left[|S_2|^2 + |S_1|^2\right] \quad (3)$$

$$S_{12} = \frac{1}{2}\left[|S_2|^2 - |S_1|^2\right] \quad (4)$$

Once the Mueller scattering matrix elements are found, the Stokes parameters can be calculated for any degree of polarization of the incident light. For example, if the incident light is polarized perpendicular to the scattering plane, the Stokes parameters of the scattered light can be found by substituting Equations 3-4 into [14]:

$$I_s = [S_{11} - S_{12}]I_i \quad (5)$$

$$Q_s = [S_{12} - S_{11}]I_i \qquad U_s = V_s = 0 \quad (6)$$

3.2. Multi Sphere Scattering

For multi sphere scattering simulation, the main complication is how the presence of another scatter center in the neighborhood affects the radiation patterns, scattering matrix and the efficiencies. When two or more identical spheres aggregate into a cluster, the resulting composite particle is nonspherical [14]. Therefore the equations for spherical particles explained in Section 3.1 are no more applicable. Recent work has verified the practical applicability of T-matrix method to clusters of spheres [16]. Hence, T-matrix approach is used for multi-sphere scattering simulation. For spheres, all T-matrix method formulas reduce to those of the standard Mie theory [13].

The cluster consists of N_s non-intersecting spheres with each sphere located at *(x,y,z)* relative to a coordinate system fixed to the cluster. Each sphere is characterized by a size parameter and a refractive index. The scattered field from the ensemble of spheres is the superposition of the fields scattered from each of the spheres [9],

$$E_s(\lambda) = \sum_{i=1}^{N_s} E_{s,i}(\lambda) \qquad E_{s,i} = \sum_{n=1}^{N_{0,i}} \sum_{m=-n}^{n} \sum_{p=1}^{2} a^i_{mnp} N_{mnp}(r_i) \quad (7)$$

where N denotes the outgoing wave vector spherical harmonic of order n and degree m and a^i is the corresponding scatterer field expansion coefficient. p indicates the mode: 1 indicates TM while 2 indicates TE mode. The cluster T-matrix is obtained following the procedure explained in detail in [10]. The elements of the amplitude scattering matrix (See Equation 1) are expressed as [9]:

$$S_1 = \tau_{mn3-p}(\theta')(-i)^n a'^{\perp}_{mnp} \qquad S_2 = \tau_{mnp}(\theta')(-i)^{n+1} a'^{\parallel}_{mnp} \quad (8)$$

$$S_3 = \tau_{mnp}(\theta')(-i)^{n+1} a'^{\perp}_{mnp} \qquad S_4 = \tau_{mn3-p}(\theta')(-i)^n a'^{\parallel}_{mnp} \quad (9)$$

where $a'^{\parallel}_{mnp}$ and $a'^{\perp}_{mnp}$ represent the scattering coefficients [9] calculated for parallel or perpendicular incident polarization, respectively. For cluster of spheres, the relationship between incident and scattered fields is conveniently written in the matrix form given in Equation 1. The relation between incident and Stokes parameters is given in Equation 2. The elements of the Mueller scattering matrix we need in our study are described in terms of the amplitude scattering matrix elements as [14]:

$$S_{11} = \frac{1}{2}\left[|S_1|^2 + |S_2|^2 + |S_3|^2 + |S_4|^2\right] \quad (10)$$

$$S_{12} = \frac{1}{2}\left[|S_2|^2 - |S_1|^2 + |S_4|^2 - |S_3|^2\right] \quad (11)$$

The Stokes parameters can be obtained as in single sphere scattering case by substituting Equations 10-11 into Equations 5-6.

4. DECOMPOSITION METHOD AND THE EXTRACTED FEATURES

In the proposed work, the specular scattering component of the LSS image is modeled by periodic or quasiperiodic image field of point scatterers corresponding to the cells' boundaries, whereas the diffused component is modeled by an autoregressive field, which corresponds to a linear filter driven by white noise. This decomposition is consistent with the general decomposition of regular stochastic fields into predictable (the specular field) component and unpredictable (the diffused field) component, known in the literature of stochastic processes as the WOLD decomposition which is used to decompose the scattered or reflection signal into its two components: diffuse and coherent [3]. The decomposition of signal is achieved by the continuous wavelet transform (CWT) and was thoroughly described and tested on simulated RF data in [11] [12]. The details of the feature extraction algorithm is explained in [12]. Two features are extracted for the coherent component: *number of coherent scatterers,* N_c and m*ean energy of the coherent scatterers, E.* Furthermore, another two features are extracted for the diffuse component which are: *residual error variance of the diffuse component,*σ^2*,* and *Rayleigh scattering degree of the diffuse component, D.* The description of all these features are given in detail in [12].

5. DETAILS OF SIMULATED AND PHANTOM DATA AND DETECTION OF DYSPLASIA FORMATION

5.1. Details of Simulated Data and Performance Evaluation on Simulations

The simulations are based on linear array of light scattering data at a given wavelength. The results on simulated overlapping and non-overlapping A-scans (256 points) data obeying Mie Scattering have been obtained. Each intensity point (See Equation 5) on the A-scan results from a cluster of *N* multi-scattering spheres located at given positions *(x, y, z).* The cluster is assumed to be confined to a space enclosed by a cube graded from -20 to +20 μm on all three axes. Unless otherwise stated, the term 'resolution cell' refers to this volume. Since we do not deal with overlapping spheres, the sphere locations are selected to be sufficiently spaced. The spheres were of a given size and index of refraction RI. The strength of the incident light I with a fixed wavelength λ (580nm in the experiment) and the direction of the detector (θ, φ) were also control parameters in the simulator. The coordinates for the first cluster ensemble is specified and the coordinates for the remaining clusters are calculated generating the 256 cluster ensemble positions. The Mueller scattering matrix corresponding to (θ,φ)=(45°,90°) is calculated (Equations 3-4 and 10-11). The intensity is calculated from the scattering matrix using Equation 5. The output of the simulator is the scattered light received at the detector from the ensemble of the scattering spheres using Mie scattering and the Mueller scattering matrix (See Equation 2). The scattered light was calculated at various positions of the resolution cell (centroids of the cube).

We have considered resolution cells and we picked parameters (cell sizes, number, index of refraction) that are close in values to normal and crowding of epithelial nuclei that are encountered in dysplasia. To introduce randomness to Mie scattering, we slightly and randomly perturb the positions, number and index of refraction of the scattering spheres around nominal values that represent a given tissue structure, and obtain 100 independent realizations of 256 points A-scans which were deemed sufficient to obtain reliable parameter estimates and good statistical samples for testing and reliable classification. For each such realization, we compute the parameters of our decomposition, and we study their discrimination power for various cases, where the scattering parameters are different.

Two different sets of simulations are done for performance evaluation. We obtain 100 different realizations with 256 points of intensity values, for each case considered. In Simulation Set 1, four types of data are created, B1, B2, S1, and S2. B1 and B2 represents the diseased cells, while S1 and S2 represents the normal cells. For the simulation of the *normal cells*, the cluster volume is assumed to contain 9-11 cells of 5μm diameter. For the simulation of the *diseased cells*, each cluster ensemble is considered to consist of 4 cells of 12μm diameter. For B1 and S1, the index of refraction is taken as 1+3i since we deal with a highly absorptive medium than reflective. In order to analyze the performance of features for classification between cells with different index of refractions the index of refraction is doubled for B2 and S2. The descriptions of these 4 types of data are given in Table 1. We take samples that are measured every 10 μm thus producing an overlap between the 256 samples so that some of the scatterers behave as coherent scatterers as well as diffused.

Table 1. Simulation Set 1

		Cell Diameter	Number of Cells	Index of Refraction
Diseased cells	**B1**	12μm	4	1+3i
	B2	12μm	4	2+6i
Normal cells	**S1**	5μm	9-11	1+3i
	S2	5μm	9-11	2+6i

In order to see the performance of the system to discriminate between single and multi scattering Simulation Set 2 is generated for which three types of data are created, A1, A2, and A3. A1 and A2 stand for a single cell with diameter 30μm and 10μm respectively with index of refraction perturbed around 1+3i. A3 represents the multi scattering case with 10 cells of diameter 5μm. The descriptions of these 3 types of data are given in Table 2.

Table 2. Simulation Set 2

		Cell Diameter	Number of Cells	Index of Refraction
Single Sphere Scattering	**A1**	30μm	1	perturbed around 1+3i
	A2	10μm	1	perturbed around 1+3i
Multi Sphere Scattering	**A3**	5μm	10	1+3i

Using the WOLD-decomposition theorem, we calculated four features, *N*, *D*, σ^2, *E* for each realization. The classification performance of each feature is evaluated separately using a quadratic classifier. Table 3 and Table 4 show the results in terms of the area under the ROC curves for all of the features separately for Simulation Set 1 and Simulation Set 2 respectively. Each value in the table presents the area under the ROC curve for classification between pairs of data. It is interesting to note here how the *KS* distance parameter *D*, which directly relates to the number of scatterers (spheres), was able to differentiate between tissue structures with different number and sizes of scattering sphere, but was not able to do the same for the same density (B1-B2 and S1-S2) but with different index of refraction. Whereas the residual error variance, σ^2 was able to discriminate very well between all four cases as the energy of the scattered light depends both the number of scatterers as well as the index of refraction, which controls the fraction of absorbed energy as compared to scattered energy. For the case of single scattering, as expected, the number of coherent scatterers N_c as well as the KS distance *D* cannot discriminate between cases A1 and A2 since they both have one scatterer per resolution cell, but do successfully differentiate when the number of scattering spheres are different (case A1-A3 and A2-A3). As expected the energy of the coherent and diffuse scatterer parameters successfully discriminate between all cases as it depends on both the index of refraction as well as the number of scatterers per resolution cell.

Table 3. A_z values for Simulation Set 1

Feature	B1-S1	B2-S2	B1-B2	S1-S2
N_c	**0.999**	**0.918**	0.545	0.985
E	0.582	1.000	0.948	1.000
D	0.741	0.857	0.540	0.635
σ^2	**1.000**	**1.000**	**1.000**	**1.000**

Table 4. of A_z values for Simulation Set 2

Feature	A1-A3	A2-A3	A1-A2
N_c	**1.000**	**1.000**	0.591
E	**1.000**	**1.000**	**0.991**
D	**0.933**	**0.957**	0.644
σ^2	**1.000**	**1.000**	**1.000**

5.2. Details of Tissue Mimicking Phantom Data and Performance Evaluation on Phantom

In our first set of experiments, we use "phantom tissue" of polystyrene latex micro spheres (Polysciences, Inc) suspended in deionized water. The micro spheres exist in a 2.5% aqueous

suspension with an index of refraction of 1.8. We examined micro spheres at diameter sizes of 3μm, and 10μm. We use a bifurcated reflectance probe (Ocean Optics, Inc), consisting of 1 central fiber and 6 surrounding fibers, each with a core of 200μm. The central fiber was used for light delivery, whereas light collection was performed by 6 external fibers. The probe has a special 30° window at the end, which reduces specular reflectance from the sample surface. The light source is a tungsten halogen lamp (LS-1 Ocean Optics, Inc). The collection part of the fiber is connected to the high resolution spectrometer (HR4000, Ocean Optics, Inc).

We use 3 μm spheres to mimic the normal cells, and 10 μm spheres to mimic the dysplastic cells. The surface area over which the data is collected is 4500 μm by 4500 μm. 3 A-scans are taken for each phantom. The scans are taken at 0 μm, 2250 μm, and 3500 μm (the y-direction). For each A-scan, 51 points are taken 90 μm apart in the x-direction. At each point the whole spectrum (at various wavelengths) is computed. Smoothing is used as a preprocessing step for each point. After the spectrum is smoothed, the intensity values for specific wavelengths are taken. As the phantom was limited in size, a permutation method is used to generate a set of 32 A-scans from each of the 3 A-scans, resulting into a total of 99 A-scans which are used for classification. The performance of the system is calculated for each specific wavelength. The classification is reported to be best for wavelengths between 500nm up to 600nm. The classification performance between the normal versus the dysplastic mimicking cells is evaluated. This is shown in Table 5 for a diffuse plus coherent scatterer model. The performance varied from 0.924 to 0.984, which is consistent with the simulation results, and is a very strong result in discriminating between normal-mimicking cells from dysplastic-mimicking ones.

Table 5. A_z values for coherent+diffuse model for Phantom Data for various wavelengths

	500.04 nm	**550.05 nm**	**571.67 nm**	**650.00 nm**
N_c	0.551	0.587	0.517	0.575
E	0.986	0.967	0.979	0.865
D	0.733	0.530	0.718	0.616
σ^2	**0.984**	**0.979**	**0.982**	**0.924**

6. CONCLUSION

The scattered light is modeled using a SDM model based on the decomposition of the two components: coherent and diffuse. Features consistent with the tissue structure are extracted from the so-called components. The performance of features to differentiate between normal and diseased cells is evaluated. The number of coherent scatterers N_c, and the energy of the diffuse component σ^2 (the residual error variance) are found to be enormously powerful parameters for detection of epithelium dysplasia formation. Our work is extremely important in terms of detection of the dysplastic epithelium in a variety of organs. Our system is able to discriminate between different sizes of nuclei, and different index of refractions, which are characteristics of dysplasia. Our results verify the ability of our system to discriminate between the normal and diseased cells. As a conclusion, our study indicates that the detection of dysplasia can be done perfectly using N_c and σ^2 based on simulations and phantom data. For the simulated data, we report on the ability of the parameters to differentiate between the diseased and normal cells with A_z=1.000 using parameter σ^2 and A_z>0.999 using parameter N_c. For the phantom data, the performance is varied from 0.927 to 0.994 depending on the wavelength used which is consistent with the simulation results, and is a very strong result in discriminating between normal-mimicking cells from dysplastic-mimicking ones. As a future work, we are planning to use real data for scattered light and try to discriminate the dysplasia formation from the normal cells.

7. REFERENCES

[1] R. S. Cotran, S. L. Robbins, and V. Kumar, Robbins Pathological Basis of Disease. Philadelphia, PA: W. B. Saunders, 1994.

[2] G. Zonios, R. M. Cothren, J. M. Crawford, M. Fitzmaurice, R. Manoharan, J. Van Dam, M. S. Feld, Spectral Pathology, Annals of the New York Academy of Sciences, 1998.

[3] F. S. Cohen, G. Georgiou, and E. Halpern, "WOLD decomposition of the backscatter echo in ultrasound images of soft tissue organs." IEEE Trans. Ultrason., Ferroelect., Freq. Contr., vol. 44, no. 2, pp. 460-472. Mar. 1997.

[4] P. H. Kaye, "Spatial light-scattering analysis as a means of characterizing and classifying non-spherical particles," Meas. Sci. Technol., vol. 9, pp. 141–149, 1998.

[5] V. Backman, R. Gurjar, K. Badizadegan, I. Itzkan, R. Dasari, L. Perelman, and M. Feld, "Polarized light scattering spectroscopy for quantitative measurement of epithelial structures in situ," IEEE J. Sel. Topics Quantum Electron., vol. 5, no. 4, pp. 1019-1026. July/August 1999.

[6] H.C. van de Hulst, Light Scattering by Small Particles, Dover Publications, New York, 1957.

[7] L.T. Perelman, J. Wu, I. Itzkan and M.S. Feld, "Photon Migration in Turbid Media Using Path-Integrals," Physical Review Letters, vol. 72, pp. 1341-1344, 1994.

[8] A. G. Yodh and B. Chance, "Spectroscopy and imaging with diffusion light," Phys. Today, vol. 48, no. 3, pp. 34–40, 1995.

[9] D. W. Mackowski and M. I. Mishchenko, "Calculation of the T matrix and the scattering matrix for ensembles of spheres ," J. Opt. Soc. Am.A, vol. 13, no. 11, pp. 2266-2278. November 1996.

[10] D. W. Mackowski, "Calculation of total cross sections of multiple-sphere clusters," J. Opt. Soc. Am. A, vol. 11, no. 11, pp. 2851-2861. November 1994.

[11] G. Georgiou and F. S. Cohen, "Tissue characterization using the continuous wavelet transform Part I: Decomposition method," IEEE Trans. Ultrason., Ferroelect., Freq. Contr., vol. 48, no. 2. pp. 355-363, Mar. 2001.

[12] G. Georgiou. F. S. Cohen, C. W. Piccoli, F. Forsberg, and B. B. Goldberg, "Tissue characterization using the continuous wavelet transform, Part II: Application on breast RF data," IEEE Trans. Ultrason., Ferroelect., Freq. Contr., vol. 48, no. 2, pp.364-373, Mar. 2001.

[13] V. P. Maltsev, and K. A. Semyanov, Characterisation of Bio-Particles from Light Scattering. Utrecht-Boston, 2004.

[14] C. F. Bohren, and D. R. Huffman, Absorption and Scattering of Light by Small Particles. John Wiley & Sons, 1983.

[15] Collett, E. Polarized Light: Fundamentals and Applications. Marcel Dekker, New York, 1993.

[16] Wielaard, D. J. , Mishchenko, M. I., Macke, A., and Carlson, B.E. "Improved T-matrix computations for large, nonabsorbing and weakly absorbing nonspherical particles and comparison with geometric optics approximation," *Applied Optics,* 36, 4305-4313, 1997.